



---

# **Psychophysiological Indicators of Aviator Flight Performance for Operator State Monitoring**

---

Kathryn Feltman, Amanda M. Kelley, Kyle Bernhardt,  
Jared Basso, & Carmelo Morabito

**DISTRIBUTION STATEMENT A.** Approved for public release; distribution unlimited.

## **Notice**

### **Qualified Requesters**

Qualified requesters may obtain copies from the Defense Technical Information Center (DTIC), Fort Belvoir, Virginia 22060. Orders will be expedited if placed through the librarian or other person designated to request documents from DTIC.

### **Change of Address**

Organizations receiving reports from the U.S. Army Aeromedical Research Laboratory on automatic mailing lists should confirm correct address when corresponding about laboratory reports.

### **Disposition**

Destroy this document when it is no longer needed. Do not return it to the originator.

### **Disclaimer**

The views, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other official documentation. Citation of trade names in this report does not constitute an official Department of the Army endorsement or approval of the use of such commercial items.

### **Human Subject Use**

In the conduct of research involving human subjects, the investigator(s) adhered to the policies regarding the protection of human subjects as prescribed by Department of Defense Instruction 32106.02 (Protection of Human Subjects and Adherence to Ethical Standards in DoD-Supported Research) dated 8 November 2011.

### **IRB Determination and Number**

This study, USAARL 2017-008, was approved by the U.S. Army Medical Research and Development Command Institutional Review Board on 10 August 2017.

**REPORT DOCUMENTATION PAGE**

*Form Approved  
OMB No. 0704-0188*

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

**PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

<b>1. REPORT DATE (DD-MM-YYYY)</b> 10-06-2021	<b>2. REPORT TYPE</b> Final Report	<b>3. DATES COVERED (From - To)</b>
--------------------------------------------------	---------------------------------------	-------------------------------------

<b>4. TITLE AND SUBTITLE</b> Psychophysiological Indicators of Aviator Flight Performance for Operator State Monitoring	<b>5a. CONTRACT NUMBER</b>
	<b>5b. GRANT NUMBER</b>
	<b>5c. PROGRAM ELEMENT NUMBER</b> 6.2

<b>6. AUTHOR(S)</b> Feltman, K.A. <sup>1</sup> , Kelley, A.M. <sup>1</sup> , Bernhardt, K. <sup>1,2</sup> , Basso, J. <sup>1,3</sup> , Morabito, C. <sup>1</sup>	<b>5d. PROJECT NUMBER</b> MO210108
	<b>5e. TASK NUMBER</b>
	<b>5f. WORK UNIT NUMBER</b>

<b>7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)</b> U.S. Army Aeromedical Research Laboratory P.O. Box 620577 Fort Rucker, AL 36362	<b>8. PERFORMING ORGANIZATION REPORT NUMBER</b> USAARL-TECH-FR--2021-17
----------------------------------------------------------------------------------------------------------------------------------------------------	----------------------------------------------------------------------------

<b>9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)</b> Military Operational Medicine Research Program U.S. Army Medical Research and Development Command 504 Scott Street Fort Detrick, MD 21702-5012	<b>10. SPONSOR/MONITOR'S ACRONYM(S)</b> MOMRP USAMRDC
	<b>11. SPONSOR/MONITOR'S REPORT NUMBER(S)</b>

**12. DISTRIBUTION/AVAILABILITY STATEMENT**  
DISTRIBUTION STATEMENT A. Approved for public release; distribution unlimited.

**13. SUPPLEMENTARY NOTES**  
<sup>1</sup>U.S. Army Aeromedical Research Laboratory, <sup>2</sup>Oak Ridge Institute for Science and Education, <sup>3</sup>Goldbelt Frontier, LLC.

**14. ABSTRACT**  
Ongoing work by the U.S. Army Aeromedical Research Laboratory (USAARL) is examining whether psychophysiological measures can reliably detect changes in operator state. If able to identify psychophysiological measures capable of reliable operator state changes, these measures could be used in operator state monitoring (OSM) of future aircraft. The present study aimed to extend previous studies completed at USAARL by evaluating whether the same measures that were previously shown to be indicators of workload could be replicated, and establish the degree to which they correlate with performance changes. To do so, individualized flights were created to control for individual differences in response to workload changes. Eight Army aviators completed the study. Workload was individually manipulated during a simulated flight scenario. Psychophysiological, subjective, and performance data were all collected to measure response to the workload manipulations. Paired samples t-tests and Pearson's correlations were performed to evaluate the data. Both demonstrated, similar to previous studies, that EEG metrics remain promising candidates for OSM.

**15. SUBJECT TERMS**  
psychophysiological indices, operator state monitoring, future vertical lift, aviation

<b>16. SECURITY CLASSIFICATION OF:</b>			<b>17. LIMITATION OF ABSTRACT</b> SAR	<b>18. NUMBER OF PAGES</b> 30	<b>19a. NAME OF RESPONSIBLE PERSON</b> Loraine St. Onge, PhD
<b>a. REPORT</b> UNCLAS	<b>b. ABSTRACT</b> UNCLAS	<b>c. THIS PAGE</b> UNCLAS			<b>19b. TELEPHONE NUMBER (Include area code)</b> 334-255-6906

This page is intentionally blank.

## **Acknowledgements**

The authors would like to express their gratitude to the research team at the U.S. Army Aeromedical Laboratory for their contributions to this project. Special gratitude is extended to Ms. Melody King who ensured the timely completion of the study with her diligent work in recruiting participants and managing the research team. Additional gratitude is extended to Mr. Rolf Beutler for assistance in preparing the flight scenarios, the research pilots (Captain [CPT] Justin Stewart, CPT Lance Randles) in Flight Systems Branch for aiding in the scenario design and data collection, and the simulator technicians in maintaining the functionality of the simulator for data collection.

This research was supported in part by an appointment to the Research Program at the U.S. Army Aeromedical Research Laboratory administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and the U.S. Army Medical Research and Development Command. Additionally, the San Francisco Model was provided by PLW Modelworks LLC, under a Cooperative Research and Development Agreement (W81XWH-19-0068) with the U.S. Army Aeromedical Research Laboratory.

This page is intentionally blank.

## Table of Contents

	<b>Page</b>
Acknowledgements.....	iii
Introduction.....	1
Methods.....	3
Participants.....	3
Materials .....	4
Questionnaires.....	4
Psychophysiological recording .....	5
Flight Tasks.....	6
Black Hawk Flight Simulator .....	6
Flight scenarios .....	6
Procedure .....	11
Visit One.....	11
Visit Two.....	12
Statistical analysis and quality control.....	12
Results.....	13
Visit One Data.....	13
Visit Two Data.....	13
Correlations between Performance and Subjective Ratings .....	15
Correlations between Psychophysiological and Subjective Ratings .....	16
Electroencephalograph.....	16
Electrocardiogram.....	17
Correlations between Psychophysiological and Performance .....	17
Electroencephalograph.....	17
Electrocardiogram.....	18
Discussion.....	18
Conclusions.....	20
References.....	21
Appendix A.....	25
Alaska and California Low Workload Scenarios.....	25
California High Workload Scenario .....	28
Appendix B.....	30
San Francisco High Workload Scenario.....	30

## List of Figures and Tables

Figure 1. NASA-TLX and airspeed correlation.....	16
Figure 2. Beta and ISAW Correlation. Shaded area indicates 95% CI.....	17
Figure 3. Frontal Alpha and Altitude Correlations. Shaded area indicates 95% CI.....	18
Table 1. Psychophysiological Workload Measures .....	1
Table 2. Questionnaires and Survey Instruments Descriptive Summary .....	4
Table 3. Visit One Flight Scenarios Summary .....	8
Table 4. Day 2 Flight Scenarios and Task Manipulations .....	10
Table 5. Individual Workload Manipulations for Visit 2 Flights .....	12
Table 6. Instantaneous Workload Ratings Summary Statistics .....	13

## Table of Contents (Continued)

	<b>Page</b>
Table 7. Visit Two Flight Performance Descriptive Statistics .....	14
Table 8. Visit Two Subjective Workload Ratings Descriptive Statistics .....	14
Table 9. Visit Two Psychophysiological Descriptive Statistics and Paired Samples t-tests .....	15
Table A1. Alaska and California Database Low Workload .....	25
Table A2. Alaska Database High Workload.....	27
Table A3. California Database High Workload.....	29

## Introduction

Army aviation is in the process of updating its fleet of aircraft with the Future Vertical Lift (FVL) program. This program anticipates delivery of new aircraft to Army aviators by 2032. With the updated aircraft will come updated technology that will enable longer duration flights, and safer flights in degraded visual environments (DVE) through increased levels of automation to support these capabilities. While the new technologies will undoubtedly provide Army aviators a competitive edge over adversaries, they will also place new demands on aviators' cognitive capacity. As the role of the operator shifts to that of a system manager, advanced technology, particularly automation, can increase an operator's workload, (e.g., Onnasch et al., 2014). Workload is defined here as the amount of attentional resources required to meet task demands (Teo et al., 2020). Tasks requiring greater attentional resources are thus high workload tasks, and place the operator at an increased risk for making performance mistakes (Dehais et al., 2020). In an effort to mitigate the risks imposed by increased aviator workload, ongoing efforts are focused on developing methods to monitor an operator's functional cognitive state in real-time. This would result in real-time operator state monitoring (OSM) implemented into future aircraft for the purposes of mitigating the negative effects of high workload (e.g., mishaps), amongst other states as well (e.g., fatigue, distraction).

Real-time OSM is monitoring an operator's functional or cognitive state as it changes in real-time. Cognitive states may include, but are not limited to, workload (overload, underload), distraction, and fatigue. In order to achieve OSM, there needs to be measurable and interpretable metrics collected from the operator in real-time that are indicative of changes to cognitive state. This can be done either through comparison to his or her steady state or using group-level metrics. Such metrics need to be able to differentiate cognitive states reliably in order to be useful to the system. Psychophysiological metrics are a candidate method for OSM due to being a relatively non-invasive, objective measure, with high temporal resolution (Teo et al., 2020). The detection of cognitive overload has been the focus of many recent efforts. Many metrics currently exist as candidates for a real-time OSM system aimed at identifying changes in workload. Table 1 (expanded on from Teo et al., 2020) summarizes several of the commonly measured psychophysiological indices of workload and the associated response patterns to high workload.

*Table 1.* Psychophysiological Workload Measures

<b>Psychophysiological Measures</b>	<b>Response to High Workload</b>	<b>Reference</b>
Heart rate	Increases	Wilson & Eggemeier, 1991
Heart rate variability (HRV)	Decreases	Mulder et al., 2004
Pupil diameter	Increases	Casali & Wierwille, 1983; May et al., 1990; Backs & Walrath, 1992; Coyne et al., 2017; Wright et al., 2014
Eye fixation duration	Increases	Callan, 1998; Wright et al., 2014; Schulz et al., 2011
Number of eye fixations	Increases <sup>1</sup> Decreases <sup>2</sup>	Van Orden et al., 2001 <sup>1</sup> ; Wright et al., 2014 <sup>2</sup>

Blink rate	Increases	Wright et al., 2014
Blink duration	Decreases (task dependent)	Merat et al., 2012
Theta waves (from EEG)	Increases	Hankins & Wilson, 1998; Feltman et al., 2020
Alpha waves (from EEG)	Decreases	Hankins & Wilson, 1998; Feltman et al., 2020
Beta waves (from EEG)	Increases	Kurimori & Kakizaki, 1995
Oxygen saturation (rSO <sub>2</sub> )	Increases	Sassaroli et al., 2008; De Joux et al., 2013
Electrodermal activity	Increases	De Waard, 1996; Daviaux et al., 2020
Cerebral blood flow velocity (CBFV)	Increases	Warm & Parasuraman, 2007
Respiration rate	Increases	Grassman et al., 2016

Identifying which of these metrics most reliably correlate with changes to performance and differentiate workload conditions are required as an initial step towards developing an OSM system. Two previous studies completed at the U.S. Army Aeromedical Research Laboratory examined the influence of task-domain (cognitive, visual, auditory, and physical) on psychophysiological recordings during tasks of varying workload (Feltman et al., 2020). These initial studies aimed primarily at evaluating the validity and reliability of psychophysiological indices as measures of workload, as well as whether different psychophysiological indices were related to the source domain of workload being taxed (i.e., cognitive, visual, auditory, physical). In these first two studies, the psychophysiological indices measured included: electroencephalogram (EEG) (frontal beta, theta, alpha band activity, and a ratio of beta/alpha+theta), electrocardiogram (ECG) (heart rate, heart rate variability), respiration (breaths per minute), and electrooculogram (EOG) (blink rate, blink duration). The researchers manipulated workload within each of the domains. The first study focused on basic task manipulations where each domain was isolated individually; whereas, the second study focused on flight tasks where each domain was the source of workload manipulations but completion of the flight scenarios inherently taxed multiple domains (see Feltman et al., 2020 for full study details).

The key takeaways from these studies included that multiple measures may be able to better detect workload levels, but that further work is needed to confirm and identify which measures are ideal candidates. Furthermore, evidence was found that cortical activity as measured by EEG may be more influenced by the domain in which the task occurs versus the degree of workload presented. In particular, the following metrics were the most reliable at differentiating workload: EEG metrics (frontal alpha, beta, theta, or their combination in the beta-ratio), and heart rate measured from ECG. These measures showed consistent differences between workload level across the various flight tasks and the basic cognitive tasks. Although some of the other measures also differentiated workload, they demonstrated much less consistency (respiration rate, blink rate, heart rate variability).

Despite some promising patterns found, these previous studies were limited by a few factors. One limitation was that the same participants were not used in both studies, thus limiting the application of findings from the first study to the second. Additionally, in the second study, participants had a wide range of flight hours indicative of various levels of experience, with

average time spent flying within the past year ranging from 12.50 to 480 hours ( $M = 198.37$ ,  $SD = 115.40$ ). It is possible that this variability in performance left some of the effects undetected, thus requiring further study prior to making a determination of which metric or set of metrics to focus on for incorporation into OSM.

In order to address some of the limitations of the previous two studies and work towards identifying variables to include in OSM, a third study was completed. In the first two studies, workload manipulations were at the group level such that every participant experienced the same conditions whereas in the third study, workload manipulations were tailored to the individual. This alternate approach was adopted to control for the effects of experience relative to flight hours and familiarity with various emergency procedures or flight in different weather conditions. To manipulate workload individually, participants visited the laboratory twice: the first visit to determine individual response to workload and a second visit to manipulate workload specific to that individual. During the first visit, the participant flew a set of four flight scenarios to evaluate responses to workload for designing the flight scenarios for the second visit. During the second visit, each participant flew two flight scenarios. The first flight scenario served as a baseline scenario to capture the participants' steady state. The second scenario included the individually tailored workload manipulations in an attempt to reach the participant's maximum level of workload. Data collected during flights included psychophysiological data, flight performance data, and subjective workload. The objectives of the study were twofold:

1) to establish the degree to which psychophysiological measures, previously shown to be valid indicators of workload (EEG, ECG), correlate to performance deficits during aviation maneuvers; and

2) to identify metrics needed for a combined suite of psychophysiological indices to monitor operator state in real-time while accounting for influence of workload domain (cognitive, physical, visual, auditory).

## Methods

The U.S. Army Medical Research and Development Command Office of Research Protections Institutional Review Board reviewed and approved the protocol for the study. All procedures were conducted according to institutional ethical standards. Participants provided written informed consent before participation.

### Participants

Eight male, Army-rated aviators participated in the study. The mean age was 37.25 years ( $SD = 3.33$ ). All participants completed two questionnaires for screening purposes: the Beck's Depression Inventory (BDI) and the State-Trait Anxiety Index (STAI) (Beck et al., 1996; Ercan et al., 2015). All participants scored below the published cut-off criterion suggesting all were free of depression (BDI,  $M = 0$ ,  $SD = 0$ ) and anxiety symptoms (STAI, State  $M = 29.14$ ,  $SD = 2.19$ , Trait  $M = 28.57$ ,  $SD = 3.91$ ), respectively. The mean Shipley Institute of Living Scale total score was 97.63 ( $SD = 5.95$ ). Seven participants were right-handed, one left-handed. No participants reported regular tobacco use (one did not respond) and four reported daily or weekly caffeine use. All participants had either uncorrected or corrected-to 20/20 vision, no other visual impairments, and no hearing impairments. Participants mean Body Mass Index scores was 27.56

( $SD = 3.05$ ), with two participants being treated for hypertension, and none reporting sleep or cognitive disorders. Participants reported a mean time in military career of 85.63 months ( $SD = 38.14$ ), or approximately 7 years. Additionally, participants reported a mean of 353.50 hours ( $SD = 158.54$ ) in the UH-60 helicopter over the course of their careers, and a mean of 8.36 hours ( $SD = 8.88$ ) on the flight controls within the past 90 days as well as 242.14 hours ( $SD = 138.02$ ) within the last year.

## Materials

Both subjective and objective instruments measured performance, individual differences, workload, and psychophysiological data.

### Questionnaires.

All questionnaires and survey instruments are presented in Table 2.

Table 2. Questionnaires and Survey Instruments Descriptive Summary

<b>Instrument</b>	<b>Description</b>
Demographics and personal history questionnaire	Brief questionnaire to collect basic demographic information including age, sex, rank, flight experience, ethnicity, and education history as well as personal health and wellness information including caffeine and tobacco use, current health, any current medications taken, attention-related diagnoses, handedness, vision, and body mass index.
Beck's Depression Inventory – II	The BDI-II is a commonly used 21-item, multiple-choice self-report that captures affect, cognitions, and physical symptoms of depression over the previous two weeks before completion. On this measure, higher scores indicate greater endorsement of depression symptoms (Beck et al., 1996).
Sleep Wake Questionnaire	The Sleep-Wake Questionnaire (SWQ) is a simple self-report survey that was developed in-house to collect information on the participant's prior night of sleep (e.g., time in bed the night prior to the study, wake time on the day of the study, if any naps were taken, and judging the quality of their sleep and current level of sleepiness on a 5-point Likert scale).
NASA Task Load Index	The NASA Task Load Index (NASA-TLX) (Hart & Staveland, 1988) is a questionnaire that measures participative workload. Participants rate each iteration of each task on the following categories, using a 100-point scale: mental demand, physical demand, temporal demand, performance, effort, and frustration. The NASA-TLX then provides a total workload score and scores for the six subscales.
ShIPLEY Institute of	The ShIPLEY Institute of Living Scale (SILS) (ShIPLEY &

Living Scale	Burlingame, 1941) assesses general intellectual functioning in adults and adolescents and serves as a screening tool for cognitive impairment in individuals with normal original intelligence (Matthews et al., 2011). The SILS takes a maximum of 20 minutes to complete, and yields three major summary scores: Vocabulary, Abstraction, and combined Total scores. The Vocabulary sub-scale consists of 40 multiple-choice verbal reasoning questions and primarily taps crystallized intelligence. The Abstraction subscale includes 20 series completion items of inductive reasoning that tap fluid ability.
State Trait Anxiety Index	The State-Trait Anxiety Inventory (STAI) (Spielberger & Gorsuch, 1983) is a widely used 40-item, self-report anxiety inventory employing a 4-point Likert-type scale that captures two types of anxiety: state, or event-dependent anxiety, and trait, or persistent demonstrations of anxiety as a personal characteristic. Anxiety scores on the STAI are calculated by reverse-coding select responses and then summing the total point values of the items, with higher scores indicating higher levels of anxiety for both the state and trait subscales.
Instantaneous Self-Assessment of Workload	The Instantaneous Self-Assessment of Workload (ISAW) (Brennan, 1992; Jordan, 1992) is a workload rating technique requiring participants to verbally rate workload on a scale of 1 (underutilized) to 5 (excessive). A programmed audio tone prompted participants when to respond. If participants did not rate workload within 5 seconds, the tone replayed. The tone replayed a total of three times (each 5 seconds apart). If the participant did not respond to any prompts, the rating was a missed response. The next rating occurred one minute after the initial tone.

---

## **Psychophysiological recording.**

### ***Electrocardiogram (ECG).***

The Biopac BioNomadix electrocardiogram amplifier module (ECG100) collected electrocardiogram (ECG) data. ECG activity with the Biopac BioNomadix records through single-lead electrodes placed on each of a participant's clavicles and one below the right pectoral area. Data samples at a rate of 1,000 hertz (Hz). The single-lead electrodes connect to a wireless transmitter belt worn around the torso that transmits the signals to a computer for recording and later processing. The research team recorded participant baseline data after ECG electrode placement. Baseline data recording required the participant to sit in a relaxed position for 5 minutes. During the baseline recordings, a member of the research team monitored the signal to ensure a clean signal is collected, indicating accurate placement of electrodes. New baseline recordings were made with each new placement of the ECG electrodes. Data were processed (including filtered, transformed, and artifacts removed) using the peak detection algorithm described in O'Brien et al. (2020). This algorithm was used to identify R-wave peaks and

identify the inter-beat-intervals (IBI); from the IBI, the outcome measures were extracted. Outcome measures included heart rate, measured in beats per minute, and heart rate variability, measured using the standard deviation of normal-to-normal (SDNN) beats.

### ***Electroencephalogram (EEG).***

The Advanced Brain Monitoring B-Alert X24 wireless wet electrode system recorded EEG activity. The X-24 incorporates 20 channels corresponding to scalp locations according to the International 10-20 system (frontal channels: Fp1, Fp2, F7, F3, Fz, F4, F8; central channels: C3, Cz, C4, T3, T4; parietal and occipital channels: P3, POz, Pz, P4, T5, T6, O1, O2). Once applied, the research team collected the participant's baseline EEG data. To collect baseline data, participants completed a series of three tasks provided with the B-Alert system. These tasks serve two purposes: 1) to create cognitive state metrics proprietary to the B-Alert system software, and 2) to collect EEG data during the various states induced by the tasks (engagement and drowsiness) for baseline information for future use. The B-Alert Live Software (Advanced Brain Monitoring, 2009) processed all EEG data. This software computes power spectral density (PSD) values, identifies and removes artifacts, and performs Fast Fourier Transform (FFT) on the data to calculate the amplitudes of the sinusoidal components for designated frequency bins. These procedures are described in (Advanced Brain Monitoring, 2009; Berka et al., 2004). Outcome measures for this study included theta (4-8 Hz), alpha (9-13 Hz), and beta (14-30 Hz) PSD values from four frontal channels (F7, F3, F4, F8). Additionally, a combination of the three, the beta-ratio (beta/alpha+theta) (Freeman et al., 2000), was also examined.

Additional psychophysiological data collected, but not reported here, included respiration (breaths per minute), EOG (blink rate, blink duration), and eye tracking / pupillometry (fixation durations, pupil diameter). These are not reported here as the focus was to establish the degree to which psychophysiological measures, previously shown to be valid indicators of workload (EEG, ECG), correlate to performance deficits during aviation maneuvers. These data have been archived for future use.

### **Flight Tasks.**

#### ***Black Hawk Flight Simulator.***

A full-motion Black Hawk simulator was used to create simulated flight scenarios (described below) designed to induce high and low workload conditions. The simulator consists of a simulator compartment containing a cockpit, instructor/operator station, observer station and a six-degree-of-freedom motion system. A Dell Precision laptop received information concerning changes in the aircraft/simulator state parameters at a 60-Hz capture rate. Settings in the simulator were manipulated to create the experience of high or low workload by manipulating the level of the clouds, the visibility, and flight system functionality.

#### ***Flight scenarios.***

Six total flight scenarios were developed. Four flight scenarios were developed for Visit One (two high and two low workload scenarios), and two flight scenarios were developed for Visit Two (one baseline and one high workload scenario). In-house Research Pilots (RPs) developed the flight scenarios and validated workload manipulations with three independent pilots. Participants flew all scenarios with a RP in the cockpit and a second RP acting as air

traffic control (ATC) in a nearby control room. The nearby control room displayed the cockpit views on video monitors and the audio from the cockpit.

***Visit One Scenarios.***

Scenarios for Visit One focused on identification of individual workload thresholds. As such, the Visit One flights were standard across all participants. The high workload flights increased workload in a linear fashion that resulted in non-flyable conditions. Flight durations in the high workload flights differed across participants, due to participants either crashing or the RP terminating the flight when maximum workload was reached, as indicated by multiple ISAW ratings of 5. The flight scenarios took place in Alaska and National Training Center (NTC), CA databases. A high workload and low workload scenario was designed for each location. Appendix A contains detailed descriptions of each Visit One flight scenario and are summarized in Table 3 below.

This space is intentionally blank.

Table 3. Visit One Flight Scenarios Summary

Database	Workload	Flight Tasks Summary	Workload Domains
	Low	Before Takeoff and Before Landing Checks Fuel Calculations Traffic Patterns	Auditory, Cognitive
Alaska	High	Before Takeoff and Before Landing Checks Fuel Calculations Multiple, Compounding Emergency Procedures Degrading Weather Conditions Incapable Co-pilot ATC Reporting Requirements	Auditory, Cognitive, Visual, Physical
	Low	Traffic Patterns Before Takeoff and Before Landing Checks Fuel Calculations	Auditory, Cognitive
California	High	Before Takeoff and Before Landing Checks Fuel Calculations Physically Demanding Emergency Procedure Urgent Mission Low Experience Co-pilot ATC Reporting Requirements	Auditory, Cognitive, Physical

***Visit Two Scenarios.***

Visit Two flight scenarios consisted of a basic flight route for both the baseline and high workload scenarios. The route required participants to conduct a flight around San Francisco, CA (PLW Modelworks, LLC). Participants flew the baseline flight first in order to collect steady-state physiological data, and to familiarize themselves with the route and database. High workload flight scenarios used responses from Visit One flights to introduce workload manipulations tailored to the individual.

Both flight scenarios used the same vignette. In the vignette, participants were recent tour guide hires for an aerial tour company based in California (“Gateway Getup”). During the baseline flight, participants were introduced to the tour company, including completion of a

“training” flight around San Francisco highlighting key attractions (e.g., Gateway Bridge, Alcatraz). The RP in the cockpit acted as the trainer for the tour company, with the RP in the control room acting as ATC. Upon completion of the first flight, the RP briefed participants that the next flight would be with their first customer, requiring the participant to fly solo. The RP adopted the role of the first customer, unable to assist the participant throughout the flight due to having no aviation background. This required participants to fly solo, which further contributed to increased workload. During this scenario, several emergencies took place along with the individualized workload manipulations. Summaries of both scenarios are in Table 4 below, with further details provided in Appendix B.

This space is intentionally blank.

Table 4. Day 2 Flight Scenarios and Task Manipulations

Event Order	Event	Workload Manipulations	Workload Domain	Estimated Time Point in minutes (m)
1	Takeoff	Patron begins asking questions	Audio/Cognitive	1:00m
2	En route	ATC requests Weather Report	Audio/Cognitive	3:00m
3	En route (Giants Stadium) VFR reporting point	Hydraulic pump #2 failure malfunction	Physical/Cognitive	5:00m
4	En route	Radio frequency change	Audio/Physical	5:00m
5	En route (Golden Gate Bridge) VFR reporting point	Precipitation/ Reduce visibility to 2 square miles (SM).	Visual/Physical	8:00m
6*	En route (Golden Gate Bridge) VFR reporting point	Increase Turbulence to level 5	Physical	8:00m
7	En route (Golden Gate Bridge) VFR reporting point	Left Pedal Drive Forward Malfunction	Physical/Cognitive	9:00m
8	En route (Golden Gate Bridge) VFR reporting point	Radio frequency change	Physical/Cognitive	9:30m
9	En route (Golden Gate Bridge) VFR reporting point	Squawk code change	Physical/Cognitive	9:30m
10	En route (49ers Stadium) VFR reporting point	Participant still flying with Pedal Drive EP	Physical	14:00m
11	En route (49ers Stadium) VFR reporting point	Reduce visibility to 1SM	Physical/Visual	16:00m
12	En route (49ers Stadium) VFR reporting point	Increased turbulence to 7	Physical	17:00m
13	En route (49ers Stadium) VFR reporting point	Engine #2 failure malfunction	Physical/Cognitive	18:00m
14	En route (49ers Stadium) VFR reporting point	Radio frequency change	Audio	19:00m
15	En route (Polo Fields) VFR reporting point/ Landing zone	None	Not Applicable	20:00m
16	Landing at Polo Fields	None	Physical	22:00m

Note. \*#2 Fuel Pressure Low malfunction was introduced here during the baseline workload flight. While workload minimally increased, the research pilot aided in the response to the malfunction. VFR = visual reporting point. ATC = air traffic control. EP = emergency procedure.

## Procedure

Participation in the study occurred over the course of two visits. Visit One determined individual workload thresholds, whereas Visit Two manipulated workload individually.

### Visit One.

When participants arrived at the laboratory, they provided written informed consent prior to participation. Following completion of the consent process, participants completed the demographics questionnaire, Beck Depression Inventory, Shipley Institute of Living Scale, Sleep Wake Questionnaire, and the State Trait Anxiety Inventory. Upon completion of the questionnaires, research technicians placed the psychophysiological devices on the participant and adjusted them for comfort. Baseline data were collected using the procedures described previously. Next, participants completed a cognitive task (*n*-back) the results of which are not reported here given that the purpose of this task is outside the scope of the present study.

The RP briefed the participant on the flight scenarios as well as use of the ISAW rating scale during the flight. During this intro brief, the RP alerted participants that, given the nature of the research, unrealistic situations could arise and that they would need to maintain specific performance criteria, potential for unrealistic situations. The RP assured the participant that their performance data would be kept confidential and only used for the purposes of the study. After the brief, the RP escorted the participant into the simulator and the participant completed a single flight to familiarize himself or herself with the laboratory's simulator. During the completion of the familiarization flight, the RP ensured that the participant was able to perform the maneuvers and the tasks required, and familiarized the participant with the ISAW. Upon confirmation of ability to meet study requirements, the flight scenarios for data collection commenced.

Participants completed one low and one high workload flight based in Alaska; and one low and one high workload flight based in NTC, for four flights total. Flights were counterbalanced using database (Alaska or NTC) and workload (high, low). ISAW ratings occurred every one-minute during each flight. Psychophysiological and flight performance metrics were collected continuously during each flight. Upon completion of the flights, a member of the research team removed psychophysiological recording devices and the participant left for the day. The two RPs met with the Principal Investigator and/or Associate Investigator to determine workload manipulations for the second day. To do so, they identified the workload source(s) the participant most struggled with: cognitive, physical, auditory, or visual. These were determined based on changes in performance (e.g., task shedding) and ISAW ratings. Table 5 summarizes the manipulations for each participant below. The two RPs then modified the high workload scenario for Visit Two to increase cognitive, auditory, physical, and/or visual workload. Workload manipulations included a set of pre-determined parameters for each workload domain (described in Table 5).

Table 5. Individual Workload Manipulations for Visit 2 Flights

<b>Workload Domain</b>	<b>Description</b>	<b>Pilots who Received Manipulation</b>
Auditory	Increase frequency of patron chatter throughout flight and increase frequency of radio calls with 40% ownship calls	$n = 7$
Cognitive	Increased tasks such as global positioning system (GPS) manipulation, radio frequency changes, and weather reports	$n = 3$
Physical	Start turbulence at 5 and increase to 9	$n = 2$

### **Visit Two.**

Upon return for Visit Two, a member of the research team affixed participants with the psychophysiological recording devices and collected new baseline data (ECG and respiration only). Participants completed the cognitive task (not reported here) and then trained on completion of the NASA-TLX. The RP briefed the participant on the day’s flights and then the participant completed the two flight scenarios (baseline flight, high workload flight). The ISAW occurred every one-minute during flight, while completion of the NASA-TLX was at the end of each flight. The psychophysiological and flight performance metrics were collected continuously during each flight. At the completion of the flights, the psychophysiological devices were removed and the participant was released for the day.

### **Statistical Analysis and Quality Control.**

All hand-entered data were double-checked for accuracy using a 10% random sample validation check. Prior to analyses, all electronically recorded data were inspected for any impossible values or output errors. Distributions of all performance and NASA-TLX ratings were evaluated for normality and inspected for outliers exceeding three standard deviations from the mean (no outliers were identified). Psychophysiological data were inspected for any impossible values with none found. All data (psychophysiological, performance, subjective) were aggregated across each flight, with means for each participant used as outcome measures. For the ISAW, this included aggregating the one-minute interval ratings into a single mean per participant.

To determine the efficacy of the workload manipulations, paired-samples *t*-tests were used to evaluate Day 1 ISAW outcomes, and Day 2 flight performance, ISAW, NASA-TLX, and psychophysiological outcomes. Both Pearson and Spearman’s (due to the small sample size) correlation tests were used to evaluate the degree to which psychophysiological measures correlated with performance deficits. However, only Pearson’s *r* is reported, as both gave similar results. All data were analyzed using R Studio, version 4.0.2, with the following packages used: “effect size” (Ben-Schachar et al., 2020) for effect size calculations, and “ggplot2” (Wickham, 2016) for correlation graphs. Hedge’s *g* is reported as it provides an unbiased effect size adjusted for small samples (Turner & Bernard, 2006).

## Results

The results are divided first by visit, with Visit Two data further divided into assessments of workload manipulations for each the subjective workload ratings, flight performance, and psychophysiological indices, and correlations between the psychophysiological indices, subjective workload ratings, and flight performance outcomes.

### Visit One Data

Given the individualized nature of the flights, data were aggregated (using the mean of the one-minute ratings) across the entirety of each flight for pairwise comparisons. Additionally, performance data were *not* analyzed with the paired samples *t*-tests, as performance in each flight were not comparable. Instead, we were interested namely in subjective ratings to verify the efficacy of the manipulations.

Paired samples *t*-tests confirmed successful manipulation of workload. Participants rated workload higher for each database for the high workload conditions, Alaska,  $t(7) = 5.40$ ,  $p = 0.001$ , Hedge's  $g = 1.70$ , and California,  $t(7) = 3.84$ ,  $p = 0.006$ , Hedge's  $g = 1.21$ .

Table 6. Instantaneous Workload Ratings Summary Statistics

Metric	Low Workload		High Workload	
	Mean	Std. Error	Mean	Std. Error
Alaska Database	2.38	0.22	3.10	0.24
California Database	1.90	0.16	2.76	0.27

### Visit Two Data

The data from the Visit Two flights, the individualized flights, are summarized below with descriptive statistics. Paired samples *t*-tests were completed to compare aggregated baseline performance data to aggregated high workload performance data. There were no significant differences during the en route portion of flight nor the landing portion. Both tail velocity and airspeed at touchdown showed trends of worse performance during the high workload condition, but significance was not reached, suggesting our sample was underpowered (see Table 7).

This space is intentionally blank.

Table 7. Visit Two Flight Performance Descriptive Statistics

Phase of Flight	Metric	Baseline		High Workload		Paired Samples <i>t</i> -tests			Hedge's <i>g</i>
		Mean	Std. Error	Mean	Std. Error	<i>t</i>	<i>df</i>	<i>p</i>	
En route	Root-mean-squared-deviations (RMSD)	64.04	4.04	66.86	3.98	-0.96	7	0.37	-
	Altitude % of Time Above 90 knots	28.56	6.34	36.17	5.00	-0.83	7	0.44	-
Landing	Tail Velocity Touchdown	2.46	0.40	5.62	1.20	-2.19	7	0.06	-
	Airspeed at Touchdown	3.60	0.76	19.08	6.32	-2.34	7	0.05	-

Paired samples *t*-tests of subjective ratings were also performed. There were statistically significant differences for each measure between the baseline and high workload flights. Descriptive statistics for the ISAW ratings, each of the NASA-TLX subscale ratings, and weighted total NASA-TLX ratings are presented in Table 8.

Table 8. Visit Two Subjective Workload Ratings Descriptive Statistics

Metric	Baseline		High Workload		Paired Samples <i>t</i> -tests			Hedge's <i>g</i>
	Mean	Std. Error	Mean	Std. Error	<i>t</i>	<i>df</i>	<i>p</i>	
ISAW Ratings	2.03	0.16	3.42	0.19	6.00	7	<0.001	1.88
*NASA-TLX - Mental	23.75	5.73	73.75	5.41	10.80	7	<0.001	3.39
NASA-TLX-Physical	12.50	4.12	75.00	10.00	6.65	7	<0.001	2.09
NASA-TLX-Temporal	20.00	5.26	71.25	7.83	6.45	7	<0.001	2.03
*NASA-TLX-Performance	13.75	3.75	45.00	7.50	3.47	7	0.01	1.09
NASA-TLX-Effort	23.75	6.73	81.88	5.17	8.15	7	<0.001	2.56
*NASA-TLX-Frustration	8.75	3.24	58.13	10.43	5.79	7	<0.001	1.82
NASA-TLX-Weighted Rating	19.25	4.42	73.71	6.26	10.04	7	<0.001	3.16

Note. \* One or both workload conditions did not meet the assumptions of a normal distribution. Robust *t*-tests were completed using 20% trimmed means with the “WRS2” package in R (Wilcox, 2020). Results remained significant, thus the results of the paired samples *t*-tests are

reported in the table to remain consistent with other results.

Paired samples *t*-tests were completed to compare aggregated baseline psychophysiological data to aggregated high workload psychophysiological data. Significant differences were found for EEG frontal alpha, theta and the beta-ratio metrics between the baseline and high workload flights (see Table 9). All metrics (frontal alpha, theta, beta-ratio) were higher during the baseline flight than during the high workload flight.

*Table 9. Visit Two Psychophysiological Descriptive Statistics and Paired Samples t-tests*

Source	Metric	Baseline		High Workload		Paired <i>t</i> -test			
		Mean	Std. Error	Mean	Std. Error	<i>t</i>	<i>df</i>	<i>p</i>	Hedge's <i>g</i>
EEG	Alpha	0.1592	0.002	0.1548*	0.002	2.46	7	0.04	0.77
	Beta	0.4051	0.006	0.4121	0.005	-2.11	7	0.07	-
	Theta	0.1727	0.007	0.1607*	0.005	3.49	7	0.01	1.10
	Beta Ratio	1.2303	0.0511	1.3112	0.0385	3.07	7	0.02	0.96
ECG	HR	82.62	4.93	89.29	7.01	1.34	6	0.23	-
	HRV	70.52	9.64	74.03	13.31	0.43	6	0.68	-

### Correlations between Performance and Subjective Ratings

Correlational analyses were completed to evaluate the relationships between the flight performance data and participants' subjective ratings of workload. There was one significant relationship. Airspeed at touchdown and NASA-TLX ratings were significantly correlated,  $r = -0.81$ ,  $p = 0.014$ , 95% CI (-0.97, -0.26). The relationship was such that as NASA-TLX ratings increased, airspeed at touchdown decreased (see Figure 1). None of the remaining measures had significant relationships (flight performance: altitude maintenance, airspeed maintenance, velocity at touchdown) with either the ISAW or NASA-TLX ratings.

This space is intentionally blank.

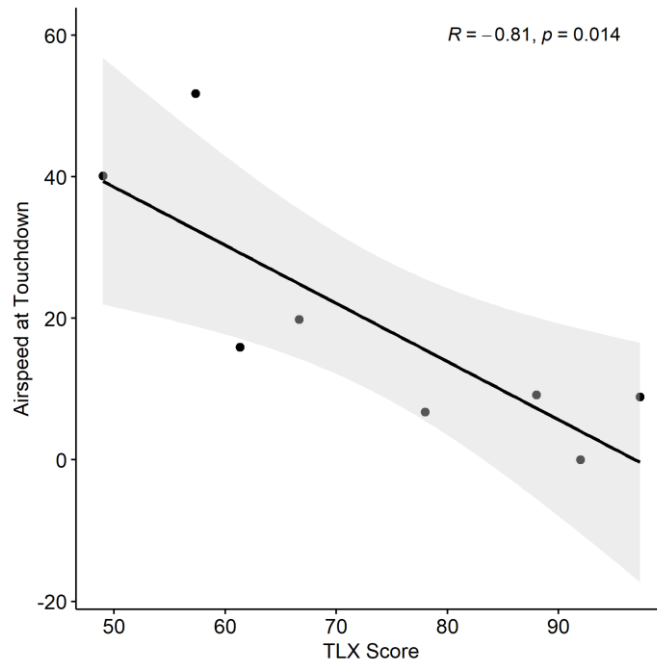


Figure 1. NASA-TLX and airspeed correlation. Shaded area indicates 95% confidence interval (CI).

### Correlations between Psychophysiological Data and Subjective Ratings

Correlational analyses were completed to evaluate the relationships between each of the psychophysiological outcomes (EEG frontal alpha, beta, theta, and beta-ratio; ECG HR and HRV) and each of the subjective ratings of workload (ISAW, NASA-TLX).

#### Electroencephalograph.

Only one significant correlation was found for this outcome metric. Participants' ISAW ratings and frontal beta activity were significantly correlated,  $r = -0.72$ ,  $p = 0.043$ , 95% CI (-0.95, -0.03). This finding indicates that as ISAW ratings increased, frontal beta activity decreased (see Figure 2).

This space is intentionally blank.

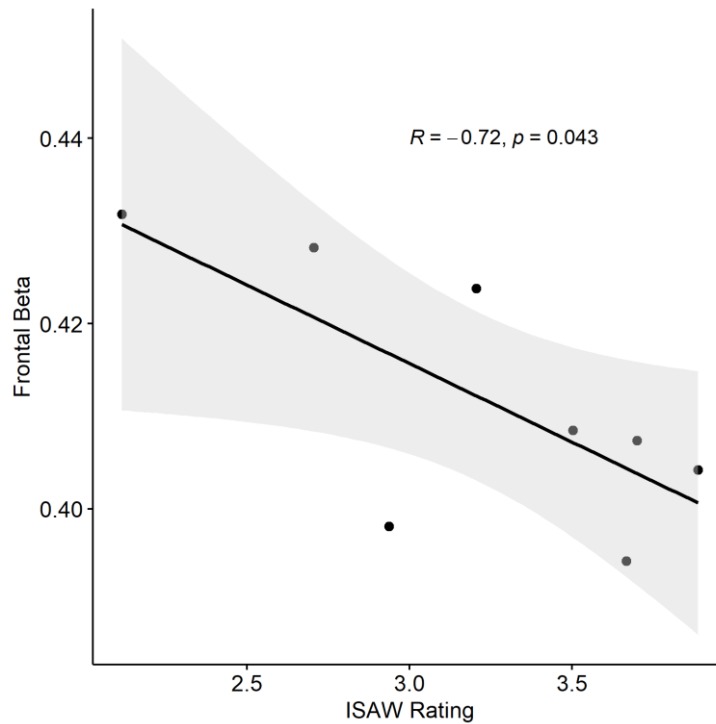


Figure 2. Beta and ISAW Correlation. Shaded area indicates 95% CI.

### **Electrocardiogram.**

There were no significant correlations between the electrocardiogram measures (HR, HRV), and either of the subjective measures (ISAW, NASA-TLX).

### **Correlations between Psychophysiological Data and Performance**

Correlational analyses were completed to evaluate the relationships between each of the each of the psychophysiological outcomes (EEG frontal alpha, beta, theta, and beta-ratio; ECG HR and HRV) and each of the flight performance outcomes (altitude maintenance, airspeed maintenance, velocity at touchdown, airspeed at touchdown).

### **Electroencephalograph.**

One significant correlation was found for this outcome metric. Participants' altitude deviations and frontal alpha activity were significantly correlated,  $r = 0.90$ ,  $p = 0.002$ , 95% CI (0.53, 0.98). The relationship is such that as altitude deviations increased, so did frontal alpha activity (see Figure 3).

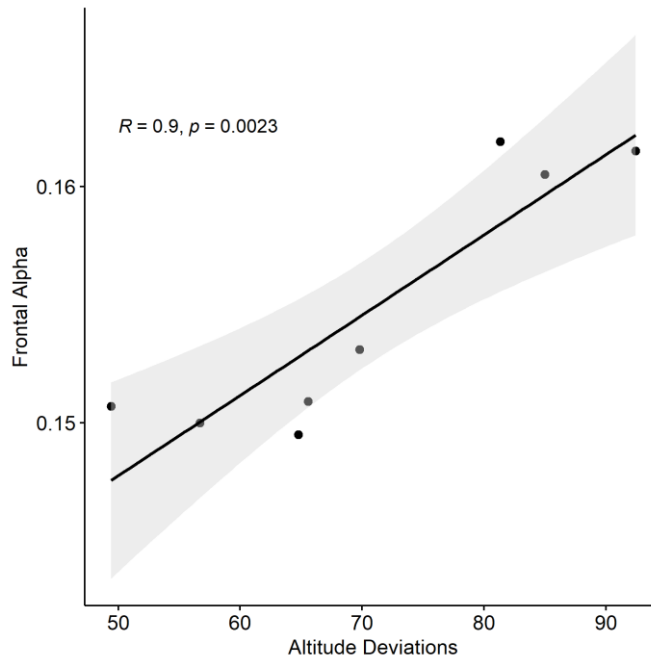


Figure 3. Frontal Alpha and Altitude Correlations. Shaded area indicates 95% CI.

### Electrocardiogram.

There were no significant correlations between the electrocardiogram measures (HR, HRV), and any of the performance outcomes (altitude maintenance, airspeed maintenance, velocity at touchdown, airspeed at touchdown).

### Discussion

This study aimed to address two objectives: 1) to establish the degree to which previously identified psychophysiological measures correlate to performance deficits during aviation maneuvers, and 2) to identify metrics needed for a combined suite of psychophysiological indices to monitor operator state. We were able to partially achieve the first objective. Specifically, frontal alpha activity again correlated with deficits in altitude performance (Feltman et al., 2020). However, in our previous study we also found a negative correlation between heart rate and airspeed, and a positive correlation between frontal alpha and airspeed. Those correlations were not replicated in this dataset. The second objective has also been partially achieved. Based on the consistency in findings from EEG measures, we can confidently recommend their consideration for inclusion in a suite of sensors to monitor operator state. However, additional sensors still require further exploration, including those collected but not reported in this study (e.g., eye tracking/pupillometry, EOG).

In the current study, we found that frontal alpha activity increases with rises in altitude deviations. This finding is consistent with past studies in simulated flight (Feltman et al., 2020). Here we found the relationship between the two yielded a Pearson's  $r$  of 0.90, whereas the previous study reported an  $r$  of 0.44. It is possible that by using the individualized workload manipulations and better controlling for confounding factors (e.g., individual differences), we were able to more accurately measure the true relationship between cortical and performance

changes. However, given the small sample size, this value should be interpreted with caution (Funder and Ozer, 2019) and requires further study to confirm.

In addition to finding the same relationship between cortical activity and altitude deviations, in Study Two, this finding occurred during the auditory domain flight. While we did not separate domains, but rather used them as a means to increase workload, seven of the eight participants received increased auditory workload as a workload manipulation. Thus, it is possible that there is a relationship with frontal alpha changes and increased altitude deviations that is dependent on auditory workload. This should be further explored with additional analyses, such as moderation models, where task domain is the moderating variable, as well as further experimental study. Moreover, all of the EEG metrics, except for frontal beta, were able to differentiate between baseline and high workload flights. These findings support further consideration of EEG metrics for OSM.

In light of the promising results, there were several limiting factors in the study to address. One limitation was the subjective nature of the workload assessments made during Visit One. Although a number of measures were taken to increase their reliability (e.g., two RPs who are also qualified instructor pilots), the process was wholly subjective and thus may have limited the validity of the manipulations used. Participants in the study also consisted of pilots with high flight hours, indicative of greater experience, which may have affected how they performed and rated the simulation, as well as any physiological changes that may have taken place. However, in an attempt to minimize these effects, the study adopted the individualized approach to workload manipulation. Regarding the physiological changes, given the experience level of the participants, it may be that EEG is more sensitive in these scenarios simply due to there being less autonomic arousal occurring. Specifically, given the “safe space” of being within a simulator versus actual aircraft, the pilots may have experienced more cognitive workload that did not influence autonomic response. Within the aircraft, the ECG data may provide additional information to support that found through EEG; however, research to-date provides mixed support.

This space is intentionally blank.

## **Conclusion**

The findings from the current study provide additional support for the continued exploration of EEG as a metric to include in OSM. In order to move towards an OSM solution, further exploration of EEG metrics is warranted. Specifically, next steps include identifying the least number of sensors needed to provide reliable information on operator state, and evaluating the best method(s) for noninvasive EEG data collection in-flight. Additionally, EEG should be paired with tasks that have potential for automation in order to begin determining how OSM can be used to trigger tasks for automation. Finally, although EEG remains the best candidate for OSM based on the series of studies conducted thus far, additional metrics, including but not limited to, ECG, eye tracking, pupillometry, and EOG, cannot yet be ruled out.

## References

- Advanced Brain Monitoring, Inc. (2009). *B-Alert Live user manual*. Carlsbad, CA.
- Backs, R. W., & Walrath, L. C. (1992). Eye movement and pupillary response indices of mental workload during visual search of symbolic displays. *Applied ergonomics*, 23(4), 243-254.
- Beck, A. T., Steer, R. A., & Brown, G. (1996). *Beck depression inventory–II*. Psychological Assessment.
- Ben-Schachar, M.S., Lüdecke, D., & Makowski, D. (2020). effectsize: Estimation of effect size indices and standardization parameters. *Journal of Open Source Software*, 5(56), 2815. <https://doi.org/10.21105/joss.02815>
- Berka, C., Levendowski, D. J., Cventinovic, M. M., Petrovic, M. M., Davis, G., Lumicao, M. N., Zivkovic, V. T., Popovich, M. V., & Olmstead, R. (2004). Real-time analysis of EEG indexes of alertness, cognition, and memory acquired with a wireless EEG headset. *International Journal of Human-Computer Interaction*, 17(2), 151-170. [https://doi.org/10.1207/s15327590ijhc1702\\_3](https://doi.org/10.1207/s15327590ijhc1702_3)
- Brennan, S. D. (1992). An experimental report on rating scale descriptor sets for the instantaneous self-assessment (ISA) recorder. DRA Technical Memorandum (CAD5) 92017, DRA Maritime Command and Control Division, Portsmouth.
- Callan, D. J. (1998, October). Eye movement relationships to excessive performance error in aviation. In *Proceedings of the Human Factors and Ergonomics Society annual meeting* (Vol. 42, No. 15, pp. 1132-1136). SAGE Publications.
- Casali, J. G., & Wierwille, W. W. (1983). A comparison of rating scale, secondary-task, physiological, and primary-task workload estimation techniques in a simulated flight task emphasizing communications load. *Human Factors*, 25(6), 623-641.
- Coyne, J. T., Foroughi, C., & Sibley, C. (2017). Pupil Diameter and Performance in a Supervisory Control Task: A Measure of Effort or Individual Differences? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 61(1), 865869. <https://doi.org/10.1177/1541931213601689>
- Daviaux, Y., Bonhomme, E., Ivers, H., de Sevin, É., Micoulaud-Franchi, J.-A., Bioulac, S., Morin, C. M., Philip, P., & Altena, E. (2020). Event-Related Electrodermal Response to Stress: Results From a Realistic Driving Simulator Scenario. *Human Factors*, 62(1), 138-151. <https://doi.org/10.1177/0018720819842779>
- Dehais, F., Lafont, A., Roy, R., & Fairclough, S. (2020). A neuroergonomics approach to mental workload, engagement, and human performance. *Frontiers in Neuroscience*, 14(268). <https://doi.org/10.3389/fnins.2020.00268>
- De Joux, N., Russell, P. N., & Helton, W. S. (2013). A functional near-infrared spectroscopy study of sustained attention to local and global target features. *Brain and Cognition*, 81, 370-375.

- de Waard, R. (1996). *The measurement of driver's mental workload*. Traffic Research Centre (now Centre for Environmental and Traffic Psychology). University of Groningen: 198.
- Ercan, I., Hafizoglu, S., Ozkaya, G., Kirli, S., Yalcintas, E., & Akaya, C. (2015). Examining cut-off values for the State-Trait Anxiety Inventory/examinando los puntajes de corte para el inventario de ansiedad estado-rasgo. *Revista Argentina de Clinica Psicologica*, 24(II), 143.
- Feltman, K. A., Bernhardt, K. A., & Kelley, A. M. (2020). Measuring the domain specificity of workload using EEG: auditory and visual domains in rotary-wing simulated flight. *Human Factors*. <https://doi.org/10.1177/0018720820928626>
- Freeman, F. G., Mikulka, P. J., Scerbo, M. W., Prinzel, L. J., & Clouatre, K. (2000). Evaluation of a psychophysiological controlled adaptive automation system, using performance on a tracking task. *Applied Psychophysiology and Biofeedback*, 25(2), 103-115.
- Funder, D. C., and Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science*, 2, 156-168. <https://doi.org/10.1177/2515245919847202>
- Grassman, M., Vlemincx, E., von Leupoldt, A., Mittelstadt, J. M., & van den Bergh, O. (2016). Respiratory changes in response to cognitive load: A systematic review. *Neural Plasticity*. <https://doi.org/10.1155/2016/8146809>
- Hankins, T. C., & Wilson, G. F. (1998). A comparison of heart rate, eye activity, EEG and subjective measures of pilot mental workload during flight. *Aviation, Space, and Environmental Medicine*, 69, 360-367.
- Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology* (Vol. 52, pp. 139-183). North-Holland.
- Jordan, C. (1992). *Experimental study of the effect of an instantaneous self assessment workload recorder on task performance* (Technical Report DRA Technical Memorandum (CAD5) 92011). DRA Maritime Command Control Division.
- Kurimori, S., & Kakizaki, T. (1995). Evaluation of work stress using psychological and physiological measures of mental activity in a paced calculating task. *Industrial health*, 33(1), 7-22.
- Matthews, T. D., Orzech, J. A., & Lassiter, K. S. (2011). Does the Shipley Institute of Living Scale measure fluid and crystallized abilities? *North American Journal of Psychology*, 13(2).
- May, J. G., Kennedy, R. S., Williams, M. C., Dunlap, W. P., & Brannan, J. R. (1990). Eye movement indices of mental workload. *Acta psychologica*, 75(1), 75-89.

- Merat, N., Jamson, A. H., Lai, F. C. H., & Carsten, O. (2012). Highly Automated Driving, Secondary Task Performance, and Driver State. *Human Factors*, 54(5), 762–771. <https://doi.org/10.1177/0018720812442087>
- Mulder, L. J. M., De Waard, D. D., & Brookhuis, K. A. (2004). Estimating mental effort using heart rate and heart rate variability. *Handbook of human factors and ergonomics methods*, pp. 666.
- O'Brien, K. J., St. Onge, P. M., & Irvin, W. D. (2020). *A peak detection algorithm for identifying R-waves in high noise electrocardiogram recordings* (Report No. USAARL-TECH-TR--2020-025). U.S. Army Aeromedical Research Laboratory.
- Onnasch, L., Wickens, C. D., Li, H., & Manzey, D. (2014). Human performance consequences of stages and levels of automation: An integrated meta-analysis. *Human factors*, 56(3), 476-488.
- Sassaroli, A., Zheng, F., Hirshfield, L. M., Girouard, A., Solovey, E. T., Jacob, R. J., & Fantini, S. (2008). Discrimination of mental workload levels in human subjects with functional near-infrared spectroscopy. *Journal of Innovative Optical Health Sciences*, 1(02), 227-237.
- Schulz, C. M., Schneider, E., Fritz, L., Vockeroth, J., Hapfelmeier, A., Wasmaier, M., & Schneider, G. (2011). Eye tracking for assessment of workload: a pilot study in an anaesthesia simulator environment. *British journal of anaesthesia*, 106(1), 44-50.
- Shibley, W. C., & Burlingame, C. C. (1941). A convenient self-administering scale for measuring intellectual impairment in psychotics. *American Journal of Psychiatry*, 97(6), 1313-1325.
- Spielberger, C. D., Gorsuch, R. L., Lushene, R.E., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the State-Trait Anxiety Inventory*. TEA Ediciones.
- Teo, G., Matthews, G., Reinerman-Jones, L., & Barber, D. (2020). Adaptive aiding with an individualized workload model based on psychophysiological measures. *Human-Intelligent Systems Integration*, 2(1), 1-15.
- Turner, H. M. III, & Bernard, R. M. (2006). Calculating and synthesizing effect sizes. *Contemporary Issues in Communication Science and Disorders*, 33, 42-55.
- Van Orden, K. F., Limbert, W., Makeig, S., & Jung, T. P. (2001). Eye activity correlates of workload during a visuospatial memory task. *Human factors*, 43(1), 111-121.
- Warm, J. S., & Parasuraman, R. (2007). Cerebral hemodynamics and vigilance. *Neuroergonomics: The brain at work*, 146-158.
- Wickham H (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag.
- Wilcox, R. M. P. (2020). Robust statistical methods in R using the WRS2 package. *Behavior Research Methods*, 52, 464-488.

- Wilson, G. F., & Eggemeier, F. T. (1991). Psychophysiological assessment of workload in multi-task environments. *Multiple-task performance*, 329360, 34.
- Wright, J. L., Quinn, S. A., Chen, J. Y. C., & Barnes, M. J. (2014). Individual differences in human agent teaming: an analysis of workload and situation awareness through eye movements. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 58(1), 1410-1414. <https://doi.org/10.1177/1541931214581294>

## Appendix A.

### Alaska and California Low Workload Scenarios

Table A1 below summarizes the flight tasks, workload manipulations, domain of workload manipulations, and time sequence of both the Alaska and California Low Workload Flight Scenario. For these scenarios, participants were briefed the following:

“You are the pilot in command (PC) and have the flight controls. We are about to conduct traffic pattern work. I will be the Instructor Pilot. The traffic pattern work will be similar to what you would experience during your annual check ride. Traffic pattern altitude will be 1000 mean sea level (MSL) and 100 knots (kts) on the downwind. 800 MSL and 80 kts on base. Your call sign in R20474.”

Weather conditions included a clear sky (no ceiling), 10 SM visibility, and no winds. Additionally, all aircraft systems were operational.

*Table A1.* Alaska and California Database Low Workload

<b>Event Order</b>	<b>Event</b>	<b>Workload Manipulations</b>	<b>Workload Domain</b>	<b>Estimated Time Point</b>
<b>1</b>	Stationary Hover Power Check	Confirm PPC data		1:00m
<b>2</b>	Takeoff			2:00m
<b>3</b>	Traffic Pattern Turns	Radio Call	Audio	4:00m
<b>4</b>	VMC Approach			6:00m
<b>5</b>	Takeoff			8:00m
<b>6</b>	Traffic Pattern Turns	Radio Call	Audio	10:00m
<b>7</b>	Roll-On Landing			12:00m
<b>8</b>	Takeoff	Hydraulic pump #2 failure malfunction	Cognitive	14:00m
<b>9</b>	Traffic Pattern Turns	Radio Call	Audio	16:00m
<b>10</b>	VMC Approach (Automatic Flight Control System [AFCS] Degraded)			18:00m
<b>11</b>	Takeoff			20:00m
<b>12</b>	Traffic Pattern Turns	Call to perform Auto Rotation	Audio	22:00m
<b>13</b>	Auto Rotation Landing			24:00m

## **Alaska High Workload Scenario**

Table A2 below summarizes the flight tasks, workload manipulations, domain of workload manipulations, and time sequence of the Alaska High Workload Flight Scenario. This scenario focused on increasing workload in multiple domains. The participant was required to initially maintain a challenging flight profile while weather began to decrease. Workload was continuously increased through the introduction of EPs combined with the lack of “experience” from the copilot. This scenario forced the participant into the clouds, losing all visual reference. EPs were continuously introduced to challenge the participant and prevent any sense of sustained situational control. The copilot became entirely unable to assist, which forced the participant to manage all EPs, navigation, and communication on his own. Eventually, the compound of EPs forced a crash sequence due to an uncontrollable aircraft.

For this scenario, participants were briefed the following:

“You will take off from Hoonah Airport (Runway 05) and fly to landing zone (LZ) Moose and contact Crusader X-Ray, a ground call sign. Here you will pick up 6 pax and drop them off at LZ Eagle (point 11 in GPS). All points of the route will be provided in the GPS for you. Due to enemy surface-to-air missile (SAM) threat you should remain at or below 250 feet above ground level (AGL) throughout transit. Once established en route maintain a speed of 120 kts. Departure and approach speeds should be flown as you would normally. Weather at Hoonah: 20 miles visibility. Ceilings are 2400’ AGL. Winds are 280/10. Your call sign is R0474. I am PI recently graduated from flight school and signed off RL1.”

This space is intentionally blank.

Table A2. Alaska Database High Workload

Event Order	Event	Workload Manipulations	Workload Domain	Estimated Time Point
1	Takeoff			1:00m
2	En route	Visibility reduced upon reaching water	Visual	2:00m
3	En route	Visibility further reduced / precipitation added	Visual	3:00m
4	En route	Radio call (frequency change)	Cognitive	4:00m
5	En route	Oil temperature malfunction	Cognitive	6:00m
6	Approach into Confined LZ		Physical/ Cognitive	8:00m
7	Landing	Visibility reduced / winds increased	Visual/Physical	8:30m
8	Loading Pax at LZ	Participant required to calculate fuel, route, and weather considerations	Cognitive	9:30m
9	Takeoff	Challenging flight route with increasing terrain elevation #1 Engine Torque (TQ) Split.	Physical/ Cognitive	11:00m
10	En route	Terrain continues to increase forcing the pilot closer to the cloud layer. Co-pilot is unable to assist with EP and navigation	Visual/Cognitive/ Physical	13:00m
11	Inadvertent Instrument Metrologic Conditions (IMC)	Co-pilot becomes disoriented. Participant is essentially single pilot, responsible for navigation and all radio communication	Cognitive/ Auditory/ Physical	15:00m
12	Inadvertent IMC	EP initiated: Stabilator Hardover failure. ATC continues to ask for information and issues holding instructions	Cognitive/ Auditory/ Physical	17:00m
13	Inadvertent IMC	EP: #1 Engine Failure. Receives holding instructions from ATC. Co-pilot provides no assistance	Cognitive/ Auditory/ Physical	19:00m
14	Inadvertent IMC	EP: SAS is disabled. Aircraft is single engine with limited controllability	Cognitive/ Auditory/ Physical	21:00m
15	Landing/Crash	Pilot loses control of aircraft	Cognitive/ Auditory/ Physical	22:00m

## California High Workload Scenario

Table A3 below summarizes the flight tasks, workload manipulations, domain of workload manipulations, and time sequence of the California High Workload Flight Scenario. This scenario increased workload by requiring the participant to complete a challenging mission with a very physically demanding EP. The co-pilot was fairly competent but unable to assist in every task (level of assistance was based on participants' proficiency). This scenario also required the use of the Garmin 530 W GPS, which added a level of cognitive difficulty, as our participants had minimal familiarity with this piece of equipment. The scenario's mission was also a mission type that was inherently associated with a sense of urgency (9-line medical evacuation [MEDEVAC]). This created an additional element of stress. Once participants received the mission and navigated to the pick-up site (while avoiding enemy small arms fire), they were required to complete a high hover hoist mission. This was eventually met with a tail rotor imbalance EP with a loss of control. Participants were required to fly back to base and perform an emergency roll-on landing to accomplish the mission.

For this scenario, participants were briefed the following:

“You are the PC and you are on the flight controls. I am your PI fresh out of flight school and progression. This is my first flight signed off readiness level 1. Our call sign is DUSTOFF 66. We are at Shukvani Airfield in Afghanistan and we are on first up MEDEVAC coverage. We have been authorized to conduct a one-hour training flight in the local area, as long as we remain on call for potential MEDEVAC missions. We are currently at MEDEVAC parking. DEMON X-Ray is our battalion ops call sign, their frequency is 41.525. Shukvani Tower is 121.5.”

This space is intentionally blank.

Table A3. California Database High Workload

<b>Event Order</b>	<b>Event</b>	<b>Workload Manipulations</b>	<b>Workload Domain</b>	<b>Estimated Time Point</b>
1	Takeoff	Confirm PPC Data		1:00m
2	Land at Hero Pad	Urgent MEDEVAC mission	Physical	2:00m
3	Enter mission details into NAV and Radio			4:00m
4	Takeoff	Taking fire from right and left sides of the aircraft	Cognitive	6:00m
5	Contact ATC for mission update	Radio Call	Audio	8:00m
6	100 foot hover hoist medic drop-off		Physical	10:00m
7	Takeoff from hover	Tail Rotor Imbalance Malfunction	Cognitive	12:00m
8	100 foot hover hoist patient and medic pick-up	Right Pedal Drive Forward Malfunction	Physical	14:00m
9	Takeoff from hover to return to base		Physical/Cognitive	16:00m
10	Flight returning to base	T/R Quadrant Fail w/ Loss of Control Malfunction	Physical	18:00m
11	Landing		Physical	20:00m

## **Appendix B.**

### **San Francisco High Workload Scenario**

This scenario required each participant to fly “single-pilot.” This left the participant responsible for all navigation, radio communication, and aircraft control. This scenario was altered for each participant based off performance in Visit One. Based off of Visit One performance, either auditory calls were increased (from the Research Pilot in the simulator control room), the “patron” would become more of a distraction to the participant, additional tasks were increased (GPS manipulation, radio frequency changes, weather reports), and/or turbulence began at a higher level. Despite these manipulations, each participant had to fly single pilot on an assigned route in a degraded visual environment with compounding EPs. The profile was completed when the participant landed at the assigned point or crashed the aircraft.

For this scenario, participants were briefed the following:

“You are the PC for this flight. You have just been signed off as the newest pilot to the Golden Gate Get-Ups, a tour company. Today you are conducting your first guided tour. I am your first patron! Along the route, there will be 4 VFR reporting points (Giants Stadium, Golden Gate Bridge, 49ers Stadium, arrival at Polo Fields). The route is loaded into the GPS and you will be responsible for all flight related activities, as I am just a paying customer for a helicopter ride around San Francisco. Maintain 200’ AGL and below and 90 knots-indicated-airspeed.”



---

## **U.S. Army Aeromedical Research Laboratory Fort Rucker, Alabama**

---

All of USAARL's science and technical information documents are available for download from the Defense Technical Information Center.

<https://discover.dtic.mil/results/?q=USAARL>



**Army Futures Command  
U.S. Army Medical Research and Development Command**